THE CROP JOURNAL 6 (2018) 299–305

Available online at www.sciencedirect.com

ScienceDirect

The Crop Journal

Identification, development, and application of cross-species intron-spanning markers in lentil (*Lens culinaris* Medik.)☆

Debjyoti Sen Gupta[a,*,1], *Jitendra Kumar*[a,*,1], *Sunanda Gupta*[a], *Sonali Dubey*[a], *Priyanka Gupta*[a], *Narendra Pratap Singh*[b], *Gaurav Sablok*[c,1]

[a]*Division of Crop Improvement, ICAR-Institute of Pulses Research, Kanpur 208024, UP, India*
[b]*Division of Plant Biotechnology, ICAR-Institute of Pulses Research, Kanpur 208024, UP, India*
[c]*Department of Biodiversity and Molecular Ecology, Research and Innovation Centre, Fondazione Edmund Mach, Via E. Mach 1, 38010 San Michele all'Adige, Trento, Italy*

ARTICLE INFO

Article history:
Received 6 July 2017
Received in revised form 21 August 2017
Accepted 25 September 2017
Available online 24 October 2017

Keywords:
Allele diversity
Cross-species
Genotypes
Introns
Lentil
Molecular markers

ABSTRACT

Lentil (*Lens culinaris* Medik) is one of the most important food legumes in the world. The use in lentil of molecular marker-assisted breeding is limited, owing to the low availability of polymorphic markers. In the present study, we developed a set of polymorphic intron-spanning markers (ISMs) using a cross-species mapping approach. In this approach, putative unique transcripts (PUTs) of *L. culinaris* were mapped onto the *Medicago truncatula* genome, exploiting its closeness with the lentil genome. Spliced alignment of the PUTs resulted in a total of 25,717 alignments, allowing the development of 1703 ISMs. From these, a subset of 105 ISMs were synthesized and validated with a 51% amplification success rate in 32 lentil genotypes. Of these ISMs, 40 (74%) were polymorphic and generated 2–11 alleles per locus in a genetically diverse panel of 32 lentil genotypes including wild species. This set of polymorphic ISMs along with their functional annotation data will be useful in lentil breeding.

1. Introduction

Lentil (*Lens culinaris* Medik.) is one of the most important food legumes and is grown in many parts of the world including Canada, Australia, northwestern USA, Turkey, Syria, Nepal, India, and Bangladesh [1]. The world annual production is nearly 5 Mt [1]. Lentil originated in the Fertile Crescent and is reported to be one of the earliest domesticated food crops [2]. Cultivation of lentil is affected by various biotic and abiotic stresses including foliar and root diseases, high temperature, drought, soil pH (<5.4), and waterlogging. Most lentil-producing countries use conventional breeding approaches in their active breeding programs for developing high-yielding lentil cultivars with better grain quality. However, in recent years, molecular markers are being widely used for accelerating precise breeding in several crops including major pulse crops such as chickpea and pigeonpea [3]. Further, the advent of next-generation sequencing technologies has allowed the rapid genome sequencing of pulses

☆ Peer review under responsibility of Crop Science Society of China and Institute of Crop Science, CAAS.

* *Corresponding authors.*
E-mail addresses: debjyoti.gupta@icar.gov.in (D.S. Gupta), jitendra.kumar@icar.gov.in (J. Kumar).

[1] Debjyoti Sen Gupta, Jitendra Kumar and Gaurav Sablok contributed equally to this work.

https://doi.org/10.1016/j.cj.2017.09.004

such as chickpea (*Cicer arietinum*) [4] and pigeonpea (*Cajanus cajan*) [5,6]. Availability of the genome sequences in these species has permitted the development of genome-anchored maps, visualization of single nucleotide polymorphism SNPs, and identifying species-specific SNPs. In the past, limited efforts have been made toward the development of molecular marker systems such as simple sequence repeats (SSRs) and SNPs for enriching genomic resources in lentil [7–12]. The plant genomes have genes with larger intronsand spliced alignment of transcripts to the genome has revealed a large diversity in intron size. Despite being of diverse lengths, introns have been a major resource for molecular-marker development in several crop species [13–16] and have been recently leveraged to develop marker resources for legumes through the development of intron-spanning markers (ISMs), which are codominant or dominant, reproducible markers that show multi-allelic patterns [13–16]. Development of these markers is gaining importance in species for which a reference genome sequence is available. In lack of the reference genome cross-species reference genome mapping can be used to identify the intron-spanning markers. Briefly, development of intron-spanning markers can be done by performing a spliced alignment of the gene transcripts to the reference genome and developing polymerase chain reaction (PCR) primers that are anchored in conserved exons that span target introns. Previously, intron-spanning markers were developed in legumes using CSGM Designer [17], which provides algorithm-or alignment-based identification of intron spanning markers and these can be validated on a set of diverse genotypes. Similarly, in lentil, transcript sequence databases available in the public domain can be used for *de novo* assembly and identification and design of primers for the amplification of ISM regions. The objectives of this study were to (1) develop polymorphic ISMs in lentil using expressed sequence tag (EST) sequences, and (2) validate polymorphic ISM markers in a diverse panel of *Lens* genotypes including wild lentil species.

2. Materials and methods

2.1. Development of intron-spanning markers, primer design, and functional annotation

A cross-species mapping-based approach was used for developing intron-spanning markers. In this approach, a well annotated and curated reference genome of *Medicago truncatula* was used because of its close phylogenetic positioning with *L. culinaris*. The *L. culinaris* putative unique transcripts (PUTs) from PlantGDB version 187 [18] were mapped onto the *M. truncatula* genome. Prior to mapping, repeat masking of the genome was performed using RepeatMasker, available from http://www.repeatmasker.org/. The RepBase libraries available from http://www.girinst.org/repbase/ and the *L. culinaris* ESTs downloaded from Plant GDB version 187 were aligned to the genome using GeneSeqer [19], a spliced alignment tool available from http://brendelgroup.org/bioinformatics2go/GeneSeqer.php. Following alignment, intron-spanning coordinates were extracted and primers were designed for the respective coordinates using Primer3 version 1.1.4, available from http://primer3.sourceforge.net/releases.php [20].

Following the identification of intron-spanning regions, primer pairs were designed using Primer3 with parameters defined as minimum amplicon size 100 bp and maximum amplicon size 300 bp, primer size 18–27 bp, primer T_m 57–63 °C, primer GC content 30%–70%, CG clamp 0, maximum end stability 250, maximum T_m difference 2, maximum self-complementarity 6, maximum 3′ end self-complementarity 3, maximum Ns accepted 0, and maximum poly-X5. The aligned PUTs to the *Medicago* genome have been annotated and then after annotation gene ontology has been defined based on the annotation. Functional annotation and gene ontology of the intron-spanning markers were performed using BLASTx searches (E-value, 1×10^{-5}) against GenBank (http://www.ncbi.nlm.nih.gov/), UniProt (http://www.uniprot.org/), and TAIR (https://www.arabidopsis.org/) databases.

2.2. Plant materials and DNA extraction

Thirty-two *Lens* genotypes were used for genotyping with 105 primers. A diverse panel of thirty-two *Lens* genotypes consisting of *L. culinaris* released cultivars, advanced breeding lines, parents of mapping populations, and genotypes of *L. ervoides* and *L. culinaris* subsp. *orientalis* was tested to identify polymorphic markers (Table 1). DNA samples were extracted from individual plant leaf tissue when seedlings were two weeks old using the cetyltrimethylammonium bromide (CTAB) procedure [21]. The DNA concentrations of the extracted samples were recorded and were compared with after corresponding concentration with λDNA. The extracted DNA samples were diluted to a uniform concentration of 20 μg μL^{-1} for PCR amplification.

2.3. PCR amplification

One hundred and five primer pairs (Table S1) were synthesized from Imperial Life Sciences (P) Limited, Gurugram, India and used in this study. PCR reactions (in 25-μL volumes) were conducted in a G-Strom (model number GT-40319, UK) thermocycler. Each reaction contained 2.5 μL *Taq* buffer (Merck, Bangalore, India), 1.5 μL $MgCl_2$ (25 mmol L^{-1}) (Merck, Bangalore, India), 0.20 mmol L^{-1} of each dNTP (Merck, Bangalore India), 0.50 mmol L^{-1} of each primer [Imperial Life Sciences (P) Limited, Gurugram India], 0.5 U of *Taq* polymerase (Merck, Bangalore, India), and 20 ng of template DNA. Primers amplifying *Lens* DNA were validated in a set of 32 diverse *Lens* genotypes using the following PCR conditions: 94 °C for 4 min, followed by 35 cycles of 94 °C for 1 min, 58 °C for 1 min, and 72 °C for 1 min followed by a final elongation step of 72 °C for 15 min. PCR products were resolved by 10% polyacrylamide vertical gel electrophoresis (Sigma-Aldrich, New Delhi, India) and visualized by silver staining. Fragments were scored visually after staining.

2.4. Molecular data scoring and statistical analysis

Polymorphism information content (PIC) values were calculated following Botstein et al. [22]. The presence and absence of the band were scored as 1 and 0 and the binary data so obtained for all *Lens* genotypes for polymorphic markers were used to calculate a correlation matrix using Jaccard's similarity coefficient analysis [23]. The similarity coefficient was used to construct a dendrogram based on the unweighted pair group

Table 1 – Details of plant materials used.

Genotype	Species	Description/pedigree	Source	Remark
IG72632	*L. culinaris* subsp. *orientalis*	Germplasm	ICARDA, Aleppo, Syria	Collected from Turkey
IG136668	*L. ervoides*	Germplasm	ICARDA, Aleppo, Syria	Collected from Turkey
IG136655	*L. culinaris* subsp. *orientalis*	Germplasm	ICARDA, Aleppo, Syria	Collected from Turkey
IG72678	*L. ervoides*	Germplasm	ICARDA, Aleppo, Syria	Collected from Syria
IG136620	*L. ervoides*	Germplasm	ICARDA, Aleppo, Syria	Collected from Slovenia
IG136615	*L. ervoides*	Germplasm	ICARDA, Aleppo, Syria	Collected from Croatia
IG72860	*L. ervoides*	Germplasm	ICARDA, Aleppo, Syria	Collected from Jordon
IG72861	*L. ervoides*	Germplasm	ICARDA, Aleppo, Syria	Collected from Jordon
IG72636	*L. culinaris* subsp. *orientalis*	Germplasm	ICARDA, Aleppo, Syria	Collected from Syria
IG116039	*L. culinaris* subsp. *orientalis*	Germplasm	ICARDA, Aleppo, Syria	Collected from Turkey
DPL62	*L. culinaris*	JL1 × LG171	IIPR, Kanpur, India	Cultivar
IPL220	*L. culinaris*	DPL44 × DPL62 × DPL58	IIPR, Kanpur, India	Cultivar
DPL58	*L. culinaris*	PL639 × PRECOZ	IIPR, Kanpur, India	Advanced breeding line
IPL526	*L. culinaris*	DPL62 × DPL58	IIPR, Kanpur, India	Cultivar
IPL81	*L. culinaris*	K75 × PL639	IIPR, Kanpur, India	Cultivar
IPL221	*L. culinaris*	DPL44 × DPL62 × DPL58	IIPR, Kanpur, India	Advanced breeding line
JL01	*L. culinaris*	Local collection from M. P. state, India	JNKVV, Jabalpur, India	Cultivar
ILL7663	*L. culinaris*	Unknown	ICARDA, Aleppo, Syria	Exotic line
IPL316	*L. culinaris*	Sehore 74-3 × DPL58	IIPR, Kanpur, India	Cultivar
IPL325	*L. culinaris*	ILL101 × E362 × DPL62	IIPR, Kanpur, India	Breeding line
IPL219	*L. culinaris*	ILL7657 × DPL61	IIPR, Kanpur, India	Breeding line
IG3973	*L. culinaris*	Germplasm	ICARDA, Aleppo, Syria	Exotic line
IG3364	*L. culinaris*	Germplasm	ICARDA, Aleppo, Syria	Exotic line
IG3575	*L. culinaris*	Germplasm	ICARDA, Aleppo, Syria	Exotic line
IG3568	*L. culinaris*	Germplasm	ICARDA, Aleppo, Syria	Exotic line
DPL15	*L. culinaris*	PL406 × L4076	IIPR, Kanpur, India	Cultivar
IG2507	*L. culinaris*	LL3	PAU, Ludhiana, India	Landrace
IG4258	*L. culinaris*	P985	USDA, RPIP, New Delhi	Landrace
FLIP2009-55L	*L. culinaris*	ILL6783 × ILL98	ICARDA, Aleppo, Syria	Exotic line
IG3327	*L. culinaris*	P1047	USDA, RPIP, New Delhi	Landrace
IG3330	*L. culinaris*	P1050	USDA, RPIP, New Delhi	Landrace
IG3546	*L. culinaris*	LG150	PAU, Ludhiana, India	Landrace

method with arithmetic average (UPGMA) using NTSYS pc-2.21q [24] software. The data were also subjected to principal coordinate analysis (PCA) using NTSYS.

3. Results

3.1. Development and validation of ISM markers in lentil

A total of 1703 ISMs were developed in lentil using a cross-species mapping-based approach. For the identification of ISMs, 16,279 EST sequences of *L. culinaris* from were mapped onto the *M. truncatula* genome (Table 2). RepBase libraries and *L. culinaris* ESTs were aligned which resulted in a total of 25,717 GeneSeqer alignments. These were further curated to identify 1703 ISMs (Table S2). Among these, a set of 105 primer pairs were used for experimental validation, which resulted in successful amplification of 54 primer pairs (51%) on lentil genomic DNA.

Table 2 – Summary statistics from bioinformatic analysis and wet-lab validation.

Parameter	Value
Number of EST sequences used for genome mapping	16,279
Number of PUT used by GeneSeqer for intron-spanning marker development	898
Number of ISM primers designed	1703
Number of primers used for validation	105
Number of polymorphic ISM primers	40
Polymorphism percentage	70

3.2. Application of ISMs in assessing genetic diversity in *Lens* species

Fifty-four ISM primer pairs were tested for identification of polymorphic markers in a diverse panel of 32 *Lens* genotypes consisting of released cultivars, advanced breeding lines, parents of mapping populations (*L. culinaris*) and genotypes of *L. ervoides* and *L. culinaris* ssp. *orientalis*. Thus, a total of 40 ISMs were found polymorphic (Table 3, Fig. 1), so that these ISMs showed high polymorphism (74%). The number of alleles ranged from 2 to 11 with an average of 3.7 alleles for each primer pair, while the PIC ranged between 0.10 and 0.50 with an average of 0.31. The sizes of alleles varied from 52 to 390 bp.

In the present study, polymorphic ISMs were used to assess the genetic diversity among the 32 genotypes and to establish the genetic relationships among them. Ten accessions belonging to different *Lens* wild species amplified 667 alleles with an average of 17 alleles per marker, while 22 accessions belonging to cultivated species amplified 1404 alleles with an average of 35 alleles per marker. These results showed high allelic diversity among the lentil genotypes for ISMs, indicating that these markers can be used further in lentil breeding programs for several purposes. Cluster analysis based on genotyping data of 40 polymorphic ISMs clustered the 32 *Lens* genotypes into two groups (Fig. 2). Group I contained 19 genotypes which belonged mostly to cultivated and wild species of lentil (*L. culinaris* subsp. *culinaris*, *L. culinaris* subsp. *orientalis*, and *L. ervoides* genotypes). Genotypes of *L. ervoides* were clustered within group I (Fig. 2).

Table 3 – T_m, allele size, and polymorphism information content (PIC) of each polymorphic intron-spanning marker.

Marker	EST sequence	Forward primer sequence (5′–3′)	Reverse primer sequence (5′–3′)	T_m (°C)	Allele	Product size	PIC value
PUT3940_1	PUT-187a-Lens_culinaris-3940	GGCGAGCGAACGGGGACCAGCC	CGGGGCTATCACCCACTATGGCCGC	73	2	52–118	0.18
PUT13107_1	PUT-187a-Lens_culinaris-13107	AGGGCGTTCAGTCGCTGGGTGTAGA	TGGGTTCGGTCCTCCAGTGCGTGT	68	3	60–118	0.33
PUT8678_1	PUT-187a-Lens_culinaris-8678	TGGAGTTGAGTTCGCCACCAGGACTCT	CGCTCCTGACCGGCAGTGTCCCA	69	4	160–230	0.37
PUT13088_1	PUT-187a-Lens_culinaris-13088	GGCGAGCGAACGGGGACCAGCC	CGGGGCTATCACCCACTATGGCCGC	59	6	52–147	0.29
PUT9449_1	PUT-187a-Lens_culinaris-9449	TGCAACTGGCTCGGAGGCTGACCA	GCGTCCGGTCCTGTTTCTCCAGGGGT	69	5	178–373	0.39
PUT5991_1	PUT-187a-Lens_culinaris-5991	TGGAAGCACCGCCCCACGAGTCA	AGGACTTCGGGTGCAATGTAAGCCGGA	68	3	167–193	0.31
PUT11257_1	PUT-187a-Lens_culinaris-11257	AAAGGGAGATTGCCGTGGCAGGGTT	AGCAGCACATTGACTCCGCAGAAGTGG	67	3	187–284	0.31
PUT2369_1	PUT-187a-Lens_culinaris-2369	ACGCTACTGGCGCTGGTGACGAAGAG	CCACGTGCACCTTCACCTGCTCCACC	70	3	138–246	0.24
PUT10315_1	PUT-187a-Lens_culinaris-10315	GCGGCGGCAGAGAAGACCGAACA	GGCCACGAGGATCTCGAAGCCTCTCA	69	4	255–329	0.24
PUT1054_1	PUT-187a-Lens_culinaris-1054	TGAGCCTACACAGACGCGGCAGCA	TGCACGAGAAACGGCCAGTTGTCCA	67	3	195–250	0.37
PUT 8925_1	PUT-187a-Lens_culinaris-8925	AGGCACCAGAGGTGGGCTAGCAGT	TGCACCGGCACTTTCCATGCCACT	67	2	163–326	0.43
PUT737_1	PUT-187a-Lens_culinaris-737	ACCGTTGGTGACGGAGCATGATGAGCA	GGCCACATTGTTGGTGGCTTCCACCGT	69	2	116–127	0.20
PUT6817_1	PUT-187a-Lens_culinaris-6817	AGAAGGACTTCTGCGCCCTCTTGGTGA	AGGCATTCACCATCTCCGGTGGGACA	68	2	131–150	0.41
PUT2753_1	PUT-187a-Lens_culinaris-2753	TGCAAGAACAGGAGGCATTGCCACGGA	GGCATGCCTTGCTTCCGCAGAAAAGGT	68	3	100–130	0.25
PUT14981_1	PUT-187a-Lens_culinaris-14981	CGCCGGTGCATTCGCCGGTCTCC	TCCAGCGCCAAAAGCAATAGAAGCCCA	68	5	209–308	0.44
PUT4935_1	PUT-187a-Lens_culinaris-4935	TGGCATGACCACTTTCCACACCGCT	TGTCACGTGCCATGAATCCAGCACCAC	67	4	196–235	0.17
PUT4843_1	PUT-187a-Lens_culinaris-4843	TGCAGCTGAGGTTTCCCGCTTTCCA	TCGAGAAGGTGCCGATCCAGCCCA	67	3	201–250	0.16
PUT6192_1	PUT-187a-Lens_culinaris-6192	TGCAACTTGCCTGCAATTCCTCACCCA	ACTGCTGCTTCCAACCCCAGCTGGTTA	67	5	240–390	0.14
PUT7040_1	PUT-187a-Lens_culinaris-7040	AACCATCACCGCCGCTTGCAACTACT	TCGGCCCAAACCCACCGGTGACA	67	2	143–167	0.46
PUT7944_1	PUT-187a-Lens_culinaris-7944	AGCCTGCTGACAGTTCCCAGTGCCA	ACAACCACGCAGGAGCAGCCAGT	67	3	201–237	0.36
PUT11770_1	PUT-187a-Lens_culinaris-11770	GCCCACCCCAGCTGCTCTTCCGA	AGAAGGAGGAGTCCTGGGGCGCTT	69	2	179–210	0.32
PUT3437_1	PUT-187a-Lens_culinaris-3437	AGGCCGCGAGAAGGAAAGTCAGGCA	GCCGACCAGAACCTCGAGCTATTCGCA	69	2	105–123	0.41
PUT3101_1	PUT-187a-Lens_culinaris-3101	GGCAAAGCGAGAAGCTCCAAAGCTCGC	GCGTCGAGACCGGAGCGAGTAGCA	70	2	245–285	0.26
PUT8778_1	PUT-187a-Lens_culinaris-8778	AGCCATTCCGTCAGTCCATCCCCCT	GCCCGCACTTCCTCAACCTTCATCGGC	69	2	270–279	0.49
PUT10666_1	PUT-187a-Lens_culinaris-10666	TGGCGGTTGTTGGCGGGACAAAGGA	GGGCATTCACATACTCGCCCGCGGAA	69	5	158–279	0.31
PUT2810_1	PUT-187a-Lens_culinaris-2810	TGCCTTGGCTATGATTGCTGGAGGGGG	GGACCTTCTTCTTGGCTGTCCAAGGGC	70	4	203–314	0.35
PUT3418_1	PUT-187a-Lens_culinaris-3418	TGCCTCCGCCGATAACACTGTCAGGC	GCGAACGGGAGCATCGTGCTTTCCCA	70	2	179–192	0.50
PUT8282_1	PUT-187a-Lens_culinaris-8282	CCGTCGGCGTTCCTCCACCGCA	ACCCCGGCGGAGGGTAAGCGT	69	2	187–201	0.50
PUT2156_1	PUT-187a-Lens_culinaris-2156	GCCGGAACATCAAGTGCTCACCAAGGC	TCGCGCAACAGGATCCGTCACAAACA	68	4	153–303	0.25
PUT11497_1A	PUT-187a-Lens_culinaris-11497	AGGGATAACGGGATGCGTGCAGCCTT	AGTCGAGCGTTCCAAGCTCTTGGGGAT	68	4	254–291	0.28
PUT11497_1B	PUT-187a-Lens_culinaris-11497	TGCTGGGCCTGCACCAAACCGCT	TGCTGGGCCTGCACCAAACCGCT	69	4	254–291	0.39
PUT11173_1	PUT-187a-Lens_culinaris-11173	GGATCAGACGCGAAGGCCGAAACCGT	TGCGGTCCTCTCCTCGCATTTCTCACC	70	6	88–220	0.25
PUT11450_1	PUT-187a-Lens_culinaris-11450	TGCGCCCAGTGCCACACCGTCG	TGCGCCCAGTGCCACACCGTCG	70	9	114–250	0.35
PUT13763_1	PUT-187a-Lens_culinaris-13763	TGGCGCCGTTTGGTGCTGCGA	GGCCGTCTGGTCAGCACTGCCTGT	68	6	62–109	0.23
PUT12611_1A	PUT-187a-Lens_culinaris-12611	GCGTTCGGCAAGGTGCACACCGTGTTA	GCGTTCGGCAAGGTGCACACCGTGTTA	70	3	159–248	0.13
PUT12611_1C	PUT-187a-Lens_culinaris-12611	GCGTTCGGCAAGGTGCACACCGTGTTA	ACCAGGGTGTTGCTTTGCTGCCTCTGC	70	3	300–337	0.49
PUT5190_1	PUT-187a-Lens_culinaris-5190	TGGTGCTGCTGCTGGAGTTGGTGCT	TGGTGCCAAAACTTTCAACCCAGGGCA	67	5	121–292	0.39
PUT13489_1	PUT-187a-Lens_culinaris-13489	GGCGGCATTTGTTGGATGCTTGATGCC	AGCTCTCCTGCGTCCAGTTTGGCGGA	69	11	156–359	0.23
PUT2042_1	PUT-187a-Lens_culinaris-2042	CCCCCAAATCCATAGAGGGTCCTGCCC	CCCCCAAATCCATAGAGGGTCCTGCCC	70	2	199–210	0.10
PUT3749_1	PUT-187a-Lens_culinaris-37491	TCTCAGGCAACCGACCTTCTGTCAGCC	TCCCTAGCTCCACTCCCCACTGGTCA	70	4	121–255	0.21

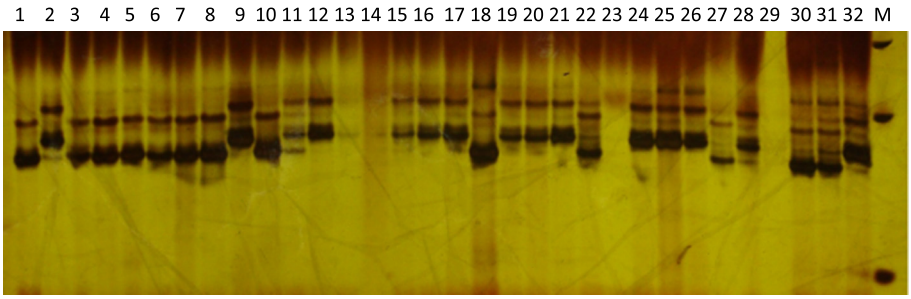

Fig. 1 – PCR amplification profiles of 32 lentil genotypes for the primer PUT 11770_1. 1, IG-72632; 2, IG-136668; 3, IG-136655; 4, IG-72678; 5, IG-136620; 6, IG-136615; 7, IG-72860; 8, IG-72861; 9, IG-72636; 10, IG-116039; 11, DPL-62; 12, IPL-220; 13, DPL-58; 14, IPL-526; 15, IPL-81; 16, IPL-221; 17, JL-01; 18, ILL-7663; 19, IPL-316; 20, IPL-325; 21, IPL-219; 22, IG3973; 23, IG3364; 24, IG3575; 25, IG3568; 26, DPL15; 27, IG2507; 28. IG4258; 29, FLIP2009-55L; 30, IG3327; 31, IG3330; 32, IG3546. Lane M contains a molecular ladder.

Another 13 accessions were clustered in group II and belonged only to *L. culinaris* subsp. *culinaris*. The first three components of PCA accounted for 13%, 9%, and 8% of total observed variation reflected as a measure of the polymorphism, respectively. In total, three PCA components accounted for 30% of total variation for the 32 genotypes.

We also studied the functional annotation of ISMs. Protein databases available in the public domain were searched for the designed ISMs. Annotations of the entire set of 1703 ISMs showed sequence similarity with legumes encoding genes, as presented in Table S3 and summarized in Fig. S1.

4. Discussion

Various molecular markers have been developed in lentil [7–12,25]. PCR-based markers such as SSRs have been the choice of plant breeders. Because the full genome sequence of lentil is not available in the public domain [11], development of polymorphic markers in lentil is slower than in other sequenced food legumes such as chickpea and pigeonpea. However, it is possible to develop intron-spanning markers (ISMs) by exploiting exon conservation to develop highly polymorphic, highly transferable, and codominant markers [26,27]. Earlier, using a cross-species mapping based approach, 91 ISMs markers were developed from pairs of highly similar genes of *M. truncatula*, a model legume, and *Glycine* max and validated on six different legume species including *M. truncatula*, *Pisum sativum*, *Lotus japonicus*, *L. filicaulis*, *Vigna radiata*, and *Phaseolus vulgaris* [28]. In this study, genomic synteny between *M. truncatula* and *Lens* was used for developing ISMs in lentil. Because ISMs can contribute toward the development of informative markers by saturating the chromosomal region of interest [26,27]. In our study, 51% of markers could be amplified on *Lens* DNA. However, in grass species, intron-spanning markers showed high (79%–95%) amplification rates on bulk DNA samples made from six forage species (*Lolium*

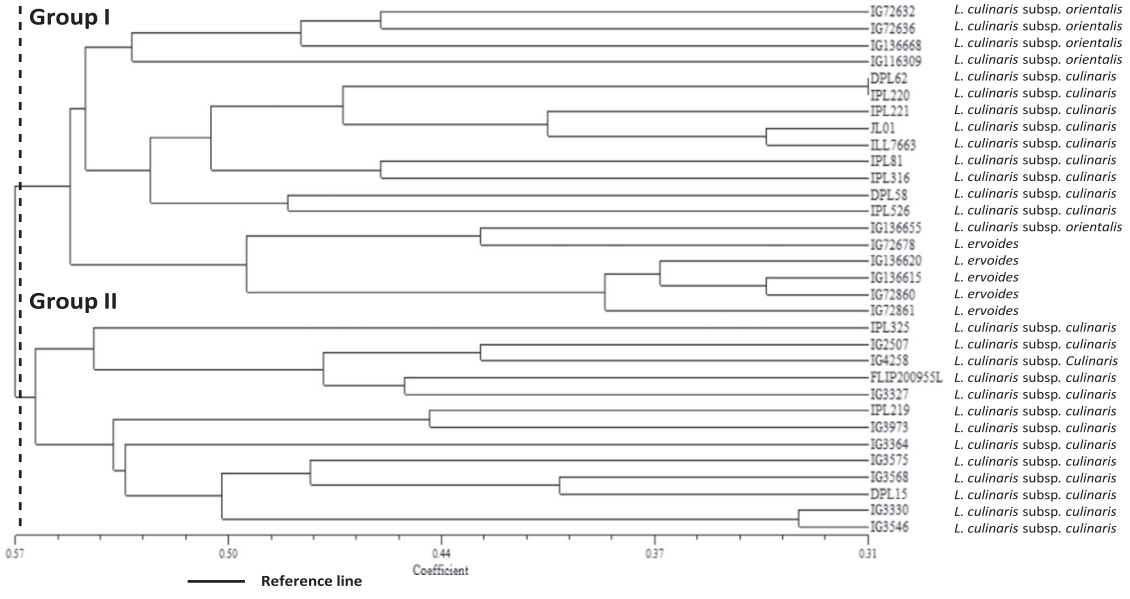

Fig. 2 – Dendrogram based on Jaccard's similarity coefficient using UPGMA clustering. The black dotted line denotes the reference line. Group I consisted of 19 genotypes (IG72632, IG72636, IG136668, IG116309, DPL62, IPL220, IPL221, JL01, ILL7663, IPL81, IPL316, DPL58, IPL526, IG136655, IG72678, IG136620, IG136615, IG72860, IG72861) and Group II consisted of 13 genotypes (IPL325, IG2507, IG4258, FLIP2009-55L, IG3327, IPL219, IG3973, IG3364, IG3575, IG3568, IG7663, IG3330, IG3546). Name of the respective *Lens* species of each genotype were given beside their names.

perenne, *L. multiflorum*, *Festuca pratensis*, *F. arundinacea*, *Phleum pretense*, and *Dactylis glomerata*) [29]. ISMs developed in the present study showed high polymorphism, 74% (40 of 54 ISMs) with an average PIC of 0.31. Similarly, ISMs were highly polymorphic in other crop species such as mustard [30] and rice [13,31]. Also, in another food legume species, pigeonpea, 55% of ISR (intron-spanning region) markers were polymorphic with an average PIC value of 0.16 and alleles amplified ranged from one to three [32]. More recently, 119,169 and 110,491 ISMs were developed from introns of *desi* and *kabuli* chickpea genes, respectively [33] and a set of 2405 ILP markers showed high polymorphism (86.2%) in 32 accessions of chickpea [33].

Our results also showed high allelic diversity among the lentil genotypes for ISMs, indicating that these markers can be useful in lentil breeding programs for several purposes. The cluster analysis based on genotyping data of 40 polymorphic ISMs clustered the 32 *Lens* genotypes into two groups. The first group showed subclustering of wild and cultivated accessions separately and further subclustering of the lentil cultivars along with a few other genotypes from ICARDA (International Center for Agricultural Research in the Dry Areas) might be due to the use of ICARDA materials in lentil breeding programs of India. The second group, comprised mostly lentil germplasm or landraces of exotic origin, indicating that they have different genetic constitution compared to cultivated accessions. Thus, the present ISMs clearly differentiated the 32 *Lens* genotypes. These ISMs will enrich current genomic resources of lentil. Further, ISMs can be used for gene expression studies for a set of candidate genes in lentil.

5. Conclusions

A set of 1703 ISMs were designed using spliced alignment of lentil EST sequences against the *Medicago* genome. A panel of 57 ISM primer pairs (polymorphic as well as monomorphic ISMs) was validated in a group of cultivated and wild lentils. These markers have been developed from gene sequences of lentil that encode several functional proteins. Therefore, these ISMs will be useful as functional or genic markers in lentil genetics and breeding.

Supplementary data for this article can be found online at https://doi.org/10.1016/j.cj.2017.09.004.

Acknowledgments

The authors thank the Indian Council of Agricultural Research, New Delhi, India for research support. This work was partially funded by the Department of Biotechnology, Government of India, New Delhi, India through a grant support (BT/PR10921/AG11/106/943/2014). Gaurav Sablok acknowledges bioinformatics server from Fondazione Edmund Mach, Italy for performing the bioinformatics analysis and development of the intron-spanning markers.

REFERENCES

[1] FAOSTAT Database, http://www.fao.org/faostat/en/#data/QC, Accessed date: 10 January 2017.

[2] S.S. Yadav, A.Z. Rizvi, M. Manohar, A.K. Verma, R. Shrestha, C. Chen, G. Bejiga, W. Chen, M. Yadav, P.N. Bahl, Lentil growers and production systems around the world, in: S.S. Yadav, L. McNeil, P.C. Stevenson (Eds.), Lentil: An Ancient Crop of Modern Times, Springer, Dordrecht, The Netherlands, 2007.

[3] R.K. Varshney, Exciting journey of 10 years from genomes to fields and markets: some success stories of genomics-assisted breeding in chickpea, pigeonpea and groundnut, Plant Sci. 242 (2016) 98–107.

[4] R.K. Varshney, C. Song, R.K. Saxena, S. Azam, S. Yu, A.G. Sharpe, S. Cannon, J. Baek, B.D. Rosen, B. Taran, T. Millan, Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement, Nat. Biotechnol. 31 (2013) 240–246.

[5] R.K. Varshney, W. Chen, Y. Li, A.K. Bharti, R.K. Saxena, J.A. Schlueter, M.T. Donoghue, S. Azam, G. Fan, A.M. Whaley, A.D. Farmer, 2012: draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers, Nat. Biotechnol. 30 (2012) 83–89.

[6] N.K. Singh, D.K. Gupta, P.K. Jayaswal, A.K. Mahato, S. Dutta, S. Singh, S. Bhutani, V. Dogra, B.P. Singh, G. Kumawat, J.K. Pal, The first draft of the pigeonpea genome sequence, J. Plant Biochem. Biotechnol. 21 (2012) 98–112.

[7] A. Hamwieh, S.M. Udupa, A. Sarkar, C. Jung, M. Baum, Development of new microsatellite markers and their application in the analysis of genetic diversity in lentils, Breed. Sci. 59 (2009) 77–86.

[8] S. Kaur, N.O. Cogan, L.W. Pembleton, M. Shinozuka, K.W. Savin, M. Materne, J.W. Forster, Transcriptome sequencing of lentil based on second-generation technology permits large-scale unigene assembly and SSR marker discovery, BMC Genomics 12 (2011) 1.

[9] S. Kaur, N.O. Cogan, A. Stephens, D. Noy, M. Butsch, J.W. Forster, M. Materne, EST-SNP discovery and dense genetic mapping in lentil (*Lens culinaris* Medik.) enable candidate gene selection for boron tolerance, Theor. Appl. Genet. 127 (2014) 703–713.

[10] P. Verma, N. Shah, S. Bhatia, Development of an expressed gene catalogue and molecular markers from the de novo assembly of short sequence reads of the lentil (*Lens culinaris* Medik.) transcriptome, Plant Biotechnol. J. 11 (2013) 894–905.

[11] K. Bett, L. Ramsay, C. Crystal, A.G. Sharpe, D.R. Cook, P.R. Varma, P. Chang, C.J. Coyne, R. McGee, D. Main, A. Vandenberg, LenGen: The International Lentil Genome Sequencing Project, Plant and Animal Genome XXIII Conference, January10–14, 2015 (San Diego, California USA).

[12] D. Sen Gupta, P. Cheng, G. Sablok, P. Thavarajah, C.J. Coyne, S. Kumar, M. Baum, R.J. McGee, Development of a panel of unigene-derived polymorphic EST-SSR markers in lentil using public database information, Crop J. 4 (2016) 425–433.

[13] S. Badoni, S. Das, Y.K. Sayal, S. Gopalakrishnan, A.K. Singh, A.R. Rao, P. Agarwal, S.K. Parida, A.K. Tyagi, Genome-wide generation and use of informative intron-spanning and intron-length polymorphism markers for high-throughput genetic analysis in rice, Sci. Rep. 6 (2016) 23765.

[14] P.J. Hiremath, A. Farmer, S.B. Cannon, J. Woodward, H. Kudapa, R. Tuteja, A. Kumar, A. Bhanuprakash, B. Mulaosmanovic, N. Gujaria, L. Krishnamurthy, Large-scale transcriptome analysis in chickpea (*Cicer arietinum* L.), an orphan legume crop of the semi-arid tropics of Asia and Africa, Plant Biotechnol. J. 9 (2011) 922–931.

[15] S. Choudhary, R. Gaur, S. Gupta, S. Bhatia, EST-derived genic molecular markers: development and utilization for generating an advanced transcript map of chickpea, Theor. Appl. Genet. 124 (2012) 1449–1462.

[16] N. Gujaria, A. Kumar, P. Dauthal, A. Dubey, P. Hiremath, A.B. Prakash, A. Farmer, M. Bhide, T. Shah, P.M. Gaur, H.D. Upadhyaya, Development and use of genic molecular markers (GMMs) for construction of a transcript map of chickpea (*Cicer arietinum* L.), Theor. Appl. Genet. 122 (2011) 1577–1589.

[17] J.H. Kim, C. Lee, D. Hyung, Y.J. Jo, J.S. Park, D.R. Cook, H.K. Choi, CSGM Designer: a platform for designing cross-species intron-

spanning genic markers linked with genome information of legumes, Plant Methods 11 (2015) 30.
[18] J. Duvick, A. Fu, U. Muppirala, M. Sabharwal, M.D. Wilkerson, C.J. Lawrence, B. Volker, PlantGDB: a resource for comparative plant genomics, Nucleic Acids Res. 36 (2008) D959–D965.
[19] V. Brendel, L. Xing, W. Zhu, Gene structure prediction from consensus spliced alignment of multiple ESTs matching the same genomic locus, Bioinformatics 20 (2004) 1157–1169.
[20] A. Untergasser, I. Cutcutache, T. Koressaar, J. Ye, B.C. Faircloth, M. Remm, S.G. Rozen, Primer3—new capabilities and interfaces, Nucleic Acids Res. 40 (2012) e115.
[21] J.J. Doyle, J.L. Doyle, A rapid DNA isolation procedure for small quantities of fresh leaf tissue, Phytochem. Bull. 19 (1987) 11–15.
[22] B. Botstein, R.L. White, M. Skolnick, R.W. Davis, Molecular markers in plant genome analysis, Am. J. Hum. Genet. 32 (1980) 314–331.
[23] P. Jaccard, Nouvellesrecherchessur la distribution florale, Bull. Soc. Vaud. Sci. Nat. 44 (1908) 223–270 (in French).
[24] F.J. Rohlf, NTSYS-pc Numerical Taxonomy and Multivariate Analysis System, Version 2.21q, Exeter Publ, New York, USA, 2009.
[25] E.E. Andeden, F.S. Baloch, E. Çakır, F. Toklu, H. Özkan, Development, characterization and mapping of microsatellite markers for lentil (*Lens culinaris* Medik.), Plant Breed. 134 (2015) 589–598.
[26] U.M. Quraishi, F. Murat, M. Abrouk, C. Pont, C. Confolent, F.X. Oury, J. Ward, D. Boros, K. Gebruers, J.A. Delcour, C.M. Courtin, Z. Bedo, L. Saulnier, F. Guillon, S. Balzergue, P.R. Shewry, C. Feuillet, G. Charmet, J. Salse, Combined meta-genomics analyses unravel candidate genes for the grain dietary fiber content in bread wheat (*Triticum aestivum* L.), Funct. Integr. Genomics 11 (2011) 71–83.
[27] U.M. Quraishi, M. Abrouk, F. Murat, C. Pont, S. Foucrier, G. Desmaizieres, C. Confolent, N. Riviere, G. Charmet, E. Paux, A. Murigneux, Cross-genome map based dissection of a nitrogen use efficiency ortho-metaQTL in bread wheat unravels concerted cereal genome evolution, Plant J. 65 (2011) 745–756.
[28] H.K. Choi, M.A. Luckow, J. Doyle, D.R. Cook, Development of nuclear gene-derived molecular markers linked to legume genetic maps, Mol. Genet. Genomics 276 (2006) 56–70.
[29] K.I. Tamura, T. Kiyoshi, J.I. Yonemaru, The development of highly transferable intron-spanning markers for temperate forage grasses, Mol. Breed. 30 (2012) 1–8.
[30] P. Panjabi, A. Jagannath, N.C. Bisht, K.L. Padmaja, S. Sharma, V. Gupta, A.K. Pradhan, D. Pental, 2008: comparative mapping of *Brassica juncea* and *Arabidopsis thaliana* using Intron Polymorphism (IP) markers: homoeologous relationships, diversification and evolution of the A, B and C *Brassica* genomes, BMC Genomics 9 (2008) 113.
[31] X.M. Xu, K.J. Liang, S.G. Zhang, W. Shang, Y.Y. Zhang, X.Y. Wei, B. Ke, Analysis of *Indica-Japonica* differentiation in rice parents and derived lines using ILP markers, Agric. Sci. China 8 (2009) 1409–1418.
[32] H. Kudapa, A.K. Bharti, S.B. Cannon, A.D. Farmer, B. Mulaosmanovic, R. Kramer, A. Bohra, N.T. Weeks, J.A. Crow, R. Tuteja, T. Shah, A comprehensive transcriptome assembly of pigeonpea (*Cajanus cajan* L.) using Sanger and second-generation sequencing platforms, Mol. Plant 5 (2012) 1020–1028.
[33] R. Srivastava, D. Bajaj, Y.K. Sayal, P.K. Meher, H.D. Upadhyaya, R. Kumar, S. Tripathi, C. Bharadwaj, A.R. Rao, S.K. Parida, Genome-wide development and deployment of informative intron-spanning and intron-length polymorphism markers for genomics-assisted breeding applications in chickpea, Plant Sci. 252 (2016) 374–387.